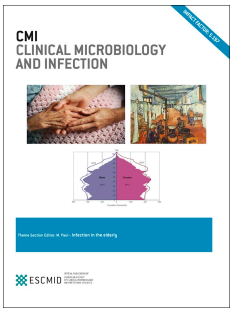
Accepted Manuscript



Detection of the NS3 Q80K polymorphism by Sanger and deep sequencing in hepatitis C virus (HCV) genotype 1a strains in the United Kingdom

Apostolos Beloukas, Simon King, Kate Childs, Athanasios Papadimitropoulos, Mark Hopkins, Mark Atkins, Kosh Agarwal, Mark Nelson, Prof Anna Maria Geretti, MD, PhD

PII: S1198-743X(15)00734-X

DOI: 10.1016/j.cmi.2015.07.017

Reference: CMI 345

To appear in: *Clinical Microbiology and Infection*

Received Date: 9 April 2015

Revised Date: 14 July 2015

Accepted Date: 19 July 2015

Please cite this article as: Beloukas A, King S, Childs K, Papadimitropoulos A, Hopkins M, Atkins M, Agarwal K, Nelson M, Geretti AM, Detection of the NS3 Q80K polymorphism by Sanger and deep sequencing in hepatitis C virus (HCV) genotype 1a strains in the United Kingdom, *Clinical Microbiology and Infection* (2015), doi: 10.1016/j.cmi.2015.07.017.

Detection of the NS3 Q80K polymorphism by Sanger and deep sequencing in hepatitis C virus (HCV) genotype 1a strains in the United Kingdom

Apostolos Beloukas[1*], Simon King[1*], Kate Childs[2], **Athanasios** Papadimitropoulos[1], Mark Hopkins[3], Mark Atkins[4], Kosh Agarwal[5], Mark Nelson[6], Anna Maria Geretti[1]

*The two authors contributed equally to this work.

[1]Institute of Infection & Global Health, University of Liverpool, Liverpool; [2]Dept. of Sexual Health, King's College Hospital NHS Foundation Trust, London; [3]Liverpool Specialist Virology Centre, Royal Liverpool University Hospital, Liverpool; [4]Dept. of Microbiology, Frimley Park Hospital NHS Foundation Trust, Frimley, Surrey; [5]Institute of Liver Studies, King's College Hospital NHS Foundation Trust, London; [6]HIV and Sexual Health Services, Chelsea and Westminster Hospital, London, United Kingdom.

Keywords: NS3, Q80K, lineage, simeprevir, protease inhibitors

Running title: Prevalence of Q80K in HCV-1a carriers in the UK

Word count: 2500

Abstract: 250

Corresponding author

Prof Anna Maria Geretti, MD, PhD

Institute of Infection & Global Health, University of Liverpool,

8 West Derby Street, L69 7BE, United Kingdom

+44 151 795 9665; geretti@liverpool.ac.uk

ABSTRACT

The Q80K polymorphism in the HCV NS3 enzyme reduces susceptibility to simeprevir and other novel protease inhibitors. The study aimed to determine the prevalence of Q80K in treatment-naïve HCV-1a carriers in the North-West (NW) and South-East (SE) of England, investigate occurrence of Q80K as a minority variant, and characterise viral phylogeny. Plasma samples from subjects that were naïve to anti-HCV therapy underwent conventional (Sanger) and deep (Illumina-Miseq, 1% interpretative cut-off) sequencing of NS3. Q80K occurred in 44/238 subjects (18.5%, 95% confidence interval 13.6%-23.4%), including 19/70 (27.1%) in the NW and 25/168 (14.9%) in the SE (p=0.0425), with no difference by HCV RNA load or HIV status. **Q80K frequencies in reads of samples that underwent Illumina sequencing were >40% in all cases. Among subjects with Q80K, 5/44 (11.4%) showed one additional major resistance-associated mutation in NS3 detected at mutant frequency above (V36L, V55A) or below (V36M) 10%.** Phylogenetic analyses identified the two recognised HCV-1a lineages with (clade I) and without (clade II) Q80K. Overall, 148/238 (62.2%) sequences occurred within regional or inter-regional clusters, **each comprising 3-20 sequences**. There was no unique clustering of English sequences relative to strains from continental Europe and North America. In conclusion, Q80K was found at high prevalence among treatment-naïve HCV-1a carriers in England, and was reliably detected by conventional sequencing, **with no increased detection by deep sequencing.** English sequences were highly interspersed with sequences from elsewhere in Europe (clade II) and North America (clade I), and their phylogeny was consistent with multiple introductions from different areas.

INTRODUCTION

The HCV NS3 protease enzyme exhibits a high degree of genetic variability and only 47% of its amino acids are conserved among circulating HCV genotypes.[1] NS3 genetic heterogeneity and the associated molecular and structural differences influence HCV susceptibility to protease inhibitors (PIs), including licensed first-generation (telaprevir, boceprevir) and second-generation (simeprevir) compounds. HCV variants with reduced PI susceptibility have been observed in treatment-naive patients.[1-10] Among these, the glutamine to lysine substitution at codon 80 (Q80K) reduces susceptibility to simeprevir *in vitro* [2, 4, 11] and was also seen to reduce virological responses to simeprevir plus pegylated interferon alpha and ribavirin (P/R) in clinical studies.[12-14] Screening for Q80K is therefore recommended before starting simeprevir.[15]

The prevalence of Q80K among PI-naïve subjects varies by HCV genotype and subtype. An analysis of global NS3 sequences deposited in GenBank reported Q80K in 42% of HCV-1a strains, and also detected a high prevalence in HCV-5 and HCV-6 strains.[1] Among HCV-1a strains, prevalence of Q80K ranges between 3% and 47% by geographical location, exceeding 40% in North America.[2, 4] Reported prevalence is 20% overall across Europe, although differing markedly by country.[7-9] Most studies to date have produced prevalence estimates using conventional (Sanger) sequencing, which allows detection of virus variants present at a frequency ≥20% within a patient's sample.

This study had three aims. Firstly, to determine the prevalence of Q80K in NS3 sequences obtained from HCV-1a carriers attending for care in two regions of England in the United Kingdom(UK). A further aim was to investigate the occurrence of Q80K as a low frequency variant by deep sequencing, and thereby both gain insights into the viral quasispecies and determine whether conventional sequencing offers sufficient sensitivity for screening simeprevir candidates for the presence of Q80K. The third aim was to determine the phylogeny of NS3 sequences in relation to publically available sequences from the rest of Europe and North America.

METHODS

Study population

The study was performed retrospectively using stored plasma samples from consecutive adults that in 2006-2014 attended for care at the Royal Liverpool University Hospital in Liverpool (North-West region) and at King's College Hospital, Chelsea & Westminster Hospital, and Charing Cross Hospital in London (South-East region). Eligible patients were infected with HCV-1a **based on the genotyping method available at the local National Health System diagnostic laboratory; the assignment was confirmed by analysis of the NS3 sequences produced in this study.** Patients were naïve to all anti-HCV therapy. HIV status and HCV RNA load were retrieved from the clinics' databases. Ethics permission was granted by the South Berkshire Regional Ethics Committee to conduct the study after removing personal identifiers from the samples.

Sanger sequencing

HCV RNA was extracted with Nuclisens easyMAG (bioMérieux, Netherlands) and cDNA synthesis was performed with the Qiagen OneStep RT-PCR Kit (Hilden, Germany) with forward primer 3278 [GGAGACCAAGMTCATCACSTGG] and reverse primer 4032 [GCTCTTRCCGCTGCCRGTGGG]. The first-round amplicon was subjected to nested PCR using the Qiagen HotStarTaq Kit with forward primer 3307 [ACACCGCSGCGTGYGGKGACAT] and reverse primer 4014 [GGRGCRTGYAGRTGGGCCAC]. The second-round 727bp amplicon (aa180 of NS2 to aa204 of NS3) was purified with the Qiagen QIAquick PCR Purification Kit and sequenced using the BigDye Terminator Cycle Sequencing Kit v3.1 on the ABI Prism 3730 Genetic Analyser (Applied Biosystems®, USA). Consensus sequences were assembled using SeqScape (v2.7) and analysed with the geno2pheno system (http://hcv.geno2pheno.org/index.php; March 2015 version) for the presence of HCV resistance-associated mutations (RAMs).

Deep sequencing

HCV RNA was extracted with QiasymphonySP (Qiagen) using the Qiagen DSPVirus/Pathogen Midi Kit and the CellFree_500_V4 protocol; cDNA was produced using the ImProm II Reverse Transcription System (Promega, USA) and random primers. NS3 was amplified using Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA) and the above primers. PCR products were purified using AMPure XP Beads (Beckman Coulter, USA) before quantification with the Quant-iT dsDNA High Sensitivity Kit (Life Technologies, USA)

on a Qubit 2.0 Fluorometer (Life Technologies). The Nextera XT Sample Preparation Kit (Illumina, USA) was used for library preparation and pooled amplicon libraries were sequenced on Illumina MiSeq using v2 reagents. Data generation and initial analysis were carried out at the Centre for Genomic Research at the University of Liverpool. Adapter sequences were trimmed from the Illumina reads using Cutadapt v1.2.1[16]; further trimming was done using Sickle v1.2 (http://github.com/najoshi/sickle) applying a minimum window quality score of 20. BAMStats (http://sourceforge.net/projects/bamstats/files/) was used to calculate the read statistics. Variants were analysed using the VirVarSeq pipeline,[17] which utilises the quality of the run and individual bases to filter poor quality bases and reduce false positive rates. A 727bp region of a HCV-1a molecular clone encoding full-length NS3 was amplified and sequenced in four replicates in order to estimate the assay error rate. The plasmid control yielded an average of 174,799 reads, with a mean coverage of 46,268 reads per nucleotide. The error rate, calculated by counting **false substitution reads** at the codon level of the molecular clone [18], was mean 0.6% (SD 0.2%) over NS3 amino acids 1-181.

Phylogenetic analysis

Reference HCV-1a sequences and all available HCV-1a NS3 sequences from North America and Europe with confirmed genotype assignment and country of origin were retrieved from the Los Alamos HCV database (hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html; accessed on 8 March 2015) **(supplementary file). All sequences, including those produced in this study (Sanger sequences and NGS consensus sequences with 1% interpretative cut-off),** were aligned pairwise against the HCV-1a reference genome H77 (accession no. NC_004102) using MEGA v.6. The alignment **was trimmed** to codon positions 1-194 of NS3 (nucleotides 3421-4014 of H77). Sequences with gaps or ambiguous nucleotides affecting ≥50% of the targeted region or where codon 80 was missing or ambiguous were removed **(n=3)**, yielding a total of 1162 NS3 sequences for analysis. Two maximum-likelihood (ML) phylogenies were estimated with and without the 80 codon, using a general-time reversible nucleotide substitution model with gamma-distributed among-site rate variation and the ML method as implemented in PHYML v.3.0.[19, 20] Phylogenetic support was evaluated using a bootstrap approach and a total of 1000 bootstrap replicates were generated using the FastTree v.2.1.7 software;[21] local branch support was calculated by the Shimodaira–Hasegawa-like (SH-like) test. Phylogenies were visualized and annotated in FigTree (http://tree.bio.ed.ac.uk). Clusters (≥3 sequences) were identified by a bootstrap support >75% and were defined as regional if

comprising sequences unique to either the NW or the SE, and inter-regional if including sequences from both regions.

Statistical analysis

Descriptive statistics were used to analyse the prevalence of mutants in the study population (as the proportion of subjects showing the mutation) and the frequency of mutants in each patient's sample (as the proportion of deep sequencing reads showing the mutation). Proportion of subjects with Q80K by geographic location and HIV status were compared by Fisher's exact test. The Mann-Whitney U test was used to compare the HCV RNA load of patients with and without Q80K. The analysis was performed using SPSS v.21.

RESULTS

Study population

Overall 238 adults (median age 44 years, with no difference by region) infected with HCV-1a were studied, including 70/238 (29.4%) in the North-West (NW) and 168/238 (70.6%) in the South-East (SE). At the time of testing, HCV RNA load was median 6.3 $\log_{10}$ IU/ml (interquartile range,IQR 5.8-6.8), without differences between the two regions. Subjects were naïve to all anti-HCV therapy. A total of 61/238 (25.6%) subjects, all from the SE, were co-infected with HIV.

Q80K prevalence by Sanger sequencing

Overall, Q80K was detected in 44/238 subjects, yielding a prevalence of 18.5% (95% confidence interval, CI 13.6%-23.4%)(Table 1). Samples from two of the 44 subjects had a mixed first base at codon 80 (Q80Q/K). **The median HCV RNA load was 6.4 (IQR 5.8-6.7) vs. 6.3 (IQR 5.8-6.8) $\log_{10}$ IU/ml in samples with vs. samples without Q80K, respectively (p=0.63)**. The prevalence of Q80K was 19/70 (27.1%) in the NW and 25/168 (14.9%) in the SE (p=0.043). When comparing results by HIV status in the SE, Q80K prevalence was 18/107 (16.8.%) in HCV mono-infected subjects and 7/61 (11.5%) in HCV/HIV co-infected subjects (p=0.379). Table 1 shows other major and minor NS3 RAMs detected by Sanger sequencing.

Q80K prevalence by deep sequencing

A total of 178/238 (74.8%) samples underwent deep sequencing, comprising 28/178 (15.7%) with and 150/178 (84.3%) without Q80K by Sanger sequencing. An average of 58,585 reads per sample was obtained (median 53,413; IQR 40,740-70,255) with an average coverage per

nucleotide of 16,107. Prevalence of Q80K was 27/178 (15.2% 95% CI 9.9%-20.5%). All 28 samples showing Q80K by Sanger sequencing also showed the mutation by deep sequencing, usually (26/28 samples) with mutant frequencies of ≥98%. Of the two samples showing Q80Q/K by Sanger sequencing, one had a Q:K ratio of 54:46, and the other showed Q, K and L at a ratio of 57:41:2. Of 150 samples lacking Q80K by Sanger sequencing, none showed the mutation by deep sequencing when applying an interpretative cut-off of ≥1% for the frequency of mutants within the reads, as usually recommended.[22] Prevalence of Q80K increased when the interpretative cut-off was lowered, and was 3/150 (2%) for Q80K occurring at a frequency between ≥0.5% and <1%, and 15/150 (10%) at a frequency ≥0.2% and <0.5%. These cut-offs however fell within the estimated error rate of the assay (0.6%, SD 0.2%).

Deep sequencing (interpretative cut-off ≥1%) identified additional NS3 RAMs that had not been observed by Sanger sequencing and which occurred at frequencies between 1% and 6% (Table 1). **Among the 44 subjects with Q80K, 4/44 (9.1%) had one additional major NS3 RAM at frequency >10% including V36L in 3/44 (6.8%) and V55A in 1/44 (2.3%). In addition 1/44 (2.3%) subjects had V36M detected only by Illumina at a frequency of 3%**.

Phylogeny of NS3 sequences

HCV-1a strains separated into the two recognised distinct lineages with and without Q80K.[23, 24] The sequences harbouring Q80K at frequency <1% did not align with the Q80K lineage (data not shown). There was no unique clustering of the study English sequences (Figure 1), rather the sequences were interspersed with sequences from the rest of Europe and North America, and the phylogeny was consistent with multiple introductions from different areas. With the English sequences, four NW, eight SE, and 11 inter-regional clusters were identified, each consisting of 3-20 sequences (Figure 2). In total, 148/238 (62.2%) sequences (49/70, 70.0% in the NW and 99/168, 58.9% in the SE) were circulating as regional or inter-regional clusters.

DISCUSSION

This study detected the NS3 polymorphism Q80K at high prevalence among treatment-naïve HCV-1a carriers attending for care in England, particularly in the North-West where prevalence reached 27.1%. Previously published data from the UK include a small cohort of 38 HIV-positive patients with acute HCV-1a infection, of which 16% had Q80K.[6] A larger study analysed 159 subjects with HCV-1a that were enrolled in Phase II/III studies of

telaprevir and simeprevir, of which 23% showed Q80K[8]; these sequences are not available for analysis however, as to date they have not been deposited in the GenBank or the Los Alamos database. These data indicate that HCV-1a strains circulating in the UK have the highest prevalence of Q80K observed in Europe.

In this study the prevalence of Q80K varied by geographical region, while showing no difference by HIV status. Prevalence rates also show marked geographical variability in the rest of Europe.[7-9] The variable prevalence has not been linked to race or ethnicity [8], but rather is consistent with the circulation of two distinct HCV-1a lineages with and without Q80K.[23, 24] The Q80K-carrying lineage is believed to have originated in the United States in the 1940s.[23] In this study, the maximum likelihood phylogenetic analysis confirmed the two HCV-1a lineages, and detected multiple introductions of the two lineages in the UK, with wide interspersing of the English strains with other European and North American strains. Most HCV-1a sequences occurred as regional or mixed regional clusters, however, suggesting that for this cohort the majority of transmission events occurred within the UK.

The deep sequencing analysis showed that Q80K mutants occurred at high frequency (>40%) in the patients' samples, allowing detection by conventional sequencing in all cases. This finding is consistent with the observation that Q80K mutants have replication capacity similar to that of wild-type virus, which allows transmissibility and persistence at high frequency within the viral quasispecies despite the absence of drug selective pressure. There were no samples showing Q80K at a frequency below the detection threshold of Sanger sequencing (≥10-20%) and above the typical ≥1% interpretative cut-off for deep sequencing, which is recommended to differentiate biologically meaningful mutations from those caused by methodological errors.[22] When applying less stringent cut-offs (0.5% or 0.2%) the proportion of samples showing Q80K by deep sequencing increased. A previous study similarly found that among 21 subjects with Q80K detected by deep sequencing, four showed the mutation at a frequency <1%.[10] The estimated error rate of the assay (0.6%) and the observation that sequences showing Q80K at a frequency <1% did not align within the Q80K lineage suggest that the detection of Q80K at frequency <1% was spurious. **Furthermore, it remains to be determined whether any Q80K variant occurring at low frequency has clinical relevance.**

Q80K reduces HCV-1a susceptibility to simeprevir by 8-11 fold *in vitro*.[4, 8, 11] In HCV-1a infection, sustained virological responses (SVR) to combination therapy with simeprevir plus P/R are reduced by the presence of baseline Q80K in both treatment-naïve subjects (from 84% to 58%) and prior relapsers (from 79% to 47%).[11-13] The impact of Q80K on the combination of simeprevir plus sofosbuvir was not immediately apparent in some reported studies.[25] **However, a phase III, open-label, single-arm study with simeprevir plus sofosbuvir in cirrhotic patients with HCV-1 recently reported that SVR rates were higher in subjects without Q80K (92%) versus those with the mutation (74%).[26]** Q80K has no effects on susceptibility to telaprevir and boceprevir [4], but may have an effect on responses to asunaprevir.[27] Q80K also confers small reductions (~3-fold) in susceptibility to faldaprevir and paritaprevir *in vitro*, without necessarily affecting response rates.[28] The Q80L and Q80R mutations were detected in some subjects, usually at low frequency. Their significance is unclear. *In vitro*, the mutations confer 2-fold and 7-fold reductions in susceptibility to simeprevir, respectively.[11] Q80R has also been observed in a small number of patients experiencing failure of simeprevir therapy, when it occurred alongside major resistance mutations at position 155 and/or position 168.[29] Other major NS3 RAMs occurred at codons 36, 54, 55, 168, and 170, which is consistent with previous observations.[1-6, 9, 30]. **A small subset of five subjects had Q80K plus one other major RAM in NS3**.

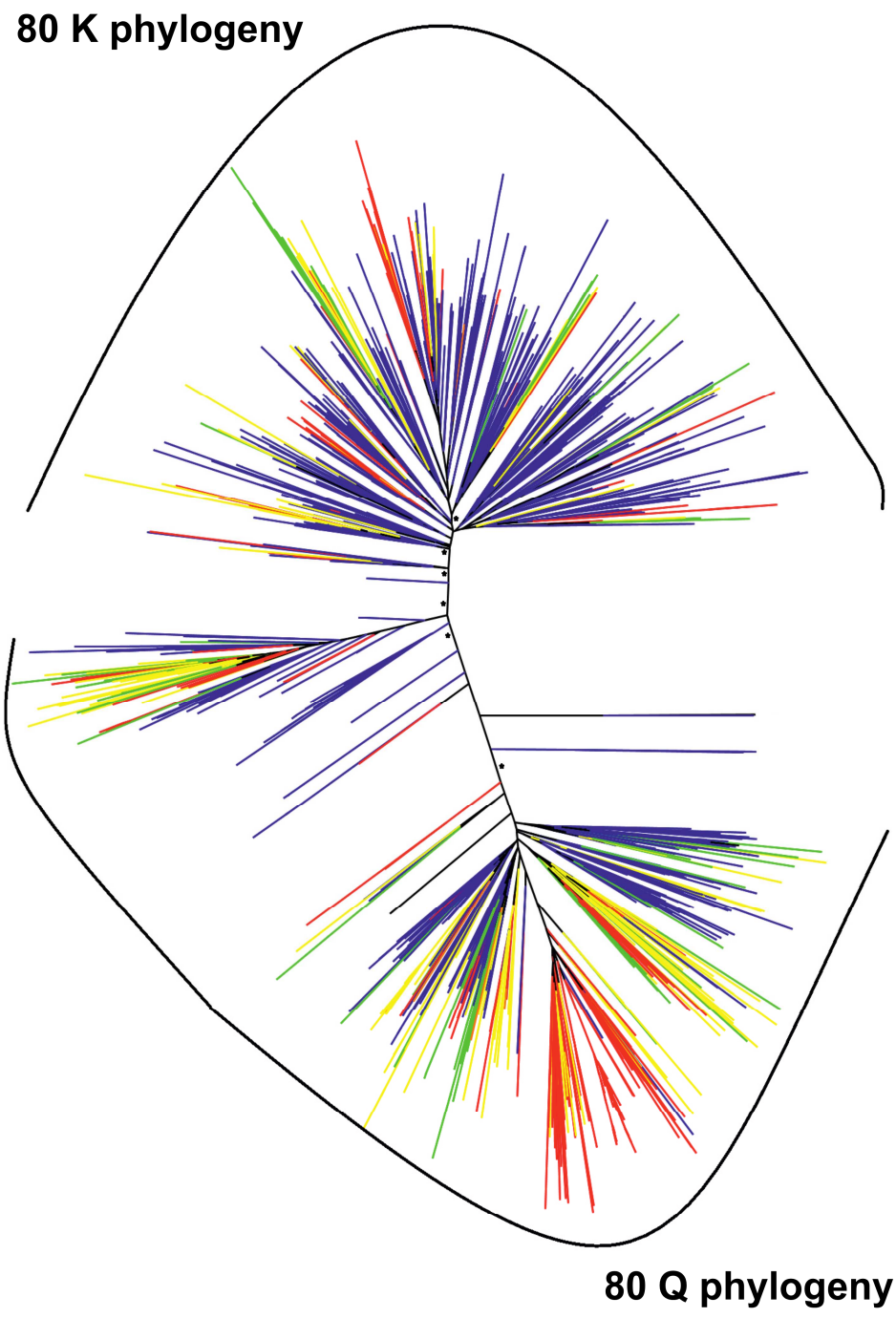
In summary, we found a high prevalence of Q80K in HCV-1a carriers with and without HIV accessing care in two regions of the UK, thus complementing estimates from elsewhere in Europe. While deep sequencing increased detection of Q80K, mutants missed by Sanger sequencing occurred at frequency ≤0.5% and detection seemed a likely technical artefact. There was evidence of a high degree of interspersing of UK sequences with sequences from elsewhere in Europe and North America, although the phylogeny indicated that most transmission events occurred in the UK. Investigation of full viral genomes will elucidate HCV-1a transmission networks and gain data to inform control strategies.

Table 1. Prevalence of NS3 resistance-associated mutations (RAMs) in treatment-naïve HCV-1a carriers **according to modality of detection (Sanger or Illumina) and frequency of the mutant as determined by Illumina**[a]

RAMs	Sanger/Illumina >10% n (%)	Illumina 1-10% n (%)
Total tested	238 (100)	178 (100)
36 M	1 (0.4)	2 (1.1)
36 L	5 (2.1)	0 (0)
54 S	9 (3.8)	0 (0)
55 A	10 (4.2)	0 (0)
80 K	44 (18.5)	0 (0)
80 L	1 (0.4)	4 (2.2)
80 R	0 (0)	3 (1.7)
117 H	0 (0)	1 (0.6)
168 E	1 (0.4)	0 (0)
170 A	1 (0.4)	0 (0)
170 T	0 (0)	2 (1.1)
174 S	97 (40.8)	10 (5.6)

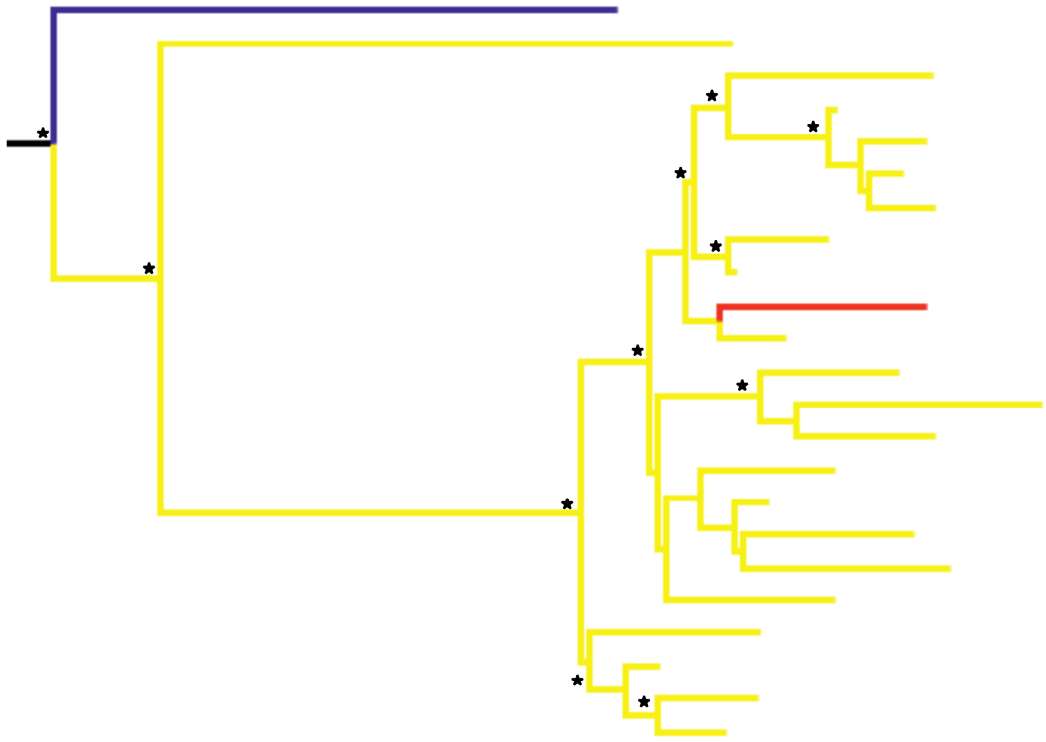
[a]RAMs were classified according to the geno2pheno mutation list (March 2015 version) with the addition of Q80L (reference 11). Bold indicates major mutations for genotype 1/1a.

277 **Figure 1**. Phylogenetic analysis of HCV-1a NS3 sequences from the United Kingdom, where
278 green indicates sequences from the North-West and yellow indicates sequences from the
279 South-East of England. **The global reference dataset was derived from Los Alamos HCV**
280 **database (March 2015; see supplementary file for accession numbers)** and comprises
281 sequences from the rest of Europe (in red) and North America (in blue). Nodes with bootstrap
282 support values ≥90% are indicated with an asterisk.
283



284 **Figure 2**. A HCV transmission cluster in the South-East of England, comprising 20 HCV-1a
285 NS3 sequences (in yellow). Sequences from the rest of Europe and North America are shown
286 in red and blue respectively. Nodes with bootstrap support values ≥90% are indicated with an
287 asterisk.

288



Acknowledgements

The work was supported by research awards from the British HIV Association and the Rosetrees Trust in the United Kingdom.

Part of this work was presented at the 13th European Meeting on HIV & Hepatitis (Barcelona, Spain; June 2015) and at the 11th International Workshop on HIV & Hepatitis Co-infection (London, UK; June 2015)

References

1 Cento V, Mirabelli C, Salpini R, et al. Hcv genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors. *PloS one*. 2012; **7**: e39652.

2 Bae A, Sun SC, Qi X, et al. Susceptibility of treatment-naive hepatitis c virus (HCV) clinical isolates to hcv protease inhibitors. *Antimicrobial Agents and Chemotherapy*. 2010; **54**: 5288-5297.

3 Bartels DJ, Sullivan JC, Zhang EZ, et al. Hepatitis c virus variants with decreased sensitivity to direct-acting antivirals (daas) were rarely observed in daa-naive patients prior to treatment. *Journal of Virology*. 2013; **87**: 1544-1553.

4 Berger KL, Triki I, Cartier M, et al. Baseline hepatitis c virus (hcv) ns3 polymorphisms and their impact on treatment response in clinical studies of the hcv ns3 protease inhibitor faldaprevir. *Antimicrobial Agents and Chemotherapy*. 2014; **58**: 698-705.

5 Kuntzen T, Timm J, Berical A, et al. Naturally occurring dominant resistance mutations to hepatitis c virus protease and polymerase inhibitors in treatment-naive patients. *Hepatology*. 2008; **48**: 1769-1778.

6 Leggewie M, Sreenu VB, Abdelrahman T, et al. Natural ns3 resistance polymorphisms occur frequently prior to treatment in hiv-positive patients with acute hepatitis c. *AIDS*. 2013; **27**: 2485-2488.

7 Morel V, Duverlie G, Brochot E. Patients eligible for treatment with simeprevir in a french center. *Journal of Clinical Virology*. 2014; **61**: 149-151.

8 Sarrazin C, Lathouwers E, Peeters M, et al. Prevalence of the hepatitis c virus ns3 polymorphism q80k in genotype 1 patients in the european region. *Antiviral Research*. 2015; **116C**: 10-16.

9 Vicenti I, Rosi A, Saladini F, et al. Naturally occurring hepatitis c virus (HCV) ns3/4a protease inhibitor resistance-related mutations in HCV genotype 1-infected subjects in italy. *Journal of Antimicrobial Chemotherapy*. 2012; **67**: 984-987.

10 Jabara CB, Hu F, Mollan KR, et al. Hepatitis c virus (HCV) NS3 sequence diversity and antiviral resistance-associated variant frequency in HCV/HIV coinfection. *Antimicrobial Agents and Chemotherapy*. 2014; **58**: 6079-6092.

11 Lenz O, Verbinnen T, Lin TI, et al. In vitro resistance profile of the hepatitis c virus ns3/4a protease inhibitor tmc435. *Antimicrobial Agents and Chemotherapy*. 2010; **54**: 1878-1887.

12 Forns X, Lawitz E, Zeuzem S, et al. Simeprevir with peginterferon and ribavirin leads to high rates of svr in patients with hcv genotype 1 who relapsed after previous therapy: A phase 3 trial. *Gastroenterology*. 2014; **146**: 1669-1679 e1663.

13 Jacobson IM, Dore GJ, Foster GR, et al. Simeprevir with pegylated interferon alfa 2a plus ribavirin in treatment-naive patients with chronic hepatitis c virus genotype 1

infection (quest-1): A phase 3, randomised, double-blind, placebo-controlled trial. *Lancet*. 2014; **384**: 403-413.

14 Manns M, Marcellin P, Poordad F, et al. Simeprevir with pegylated interferon alfa 2a or 2b plus ribavirin in treatment-naive patients with chronic hepatitis c virus genotype 1 infection (quest-2): A randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*. 2014; **384**: 414-426.

15 Simeprevir summary of product characteristics. Available at: Http://www.Ema.Europa.Eu/docs/en_gb/document_library/epar_-_product_information/human/002777/wc500167867.Pdf. (accessed 3 march 2015).

16 M. M. Cutadapt removes adapter sequences from high-throughput sequencing reads. . *EMBnet J* 2011; **17**: 10-12.

17 Verbist BM, Thys K, Reumers J, et al. Virvarseq: A low-frequency virus variant detection pipeline for illumina sequencing using adaptive base-calling accuracy filtering. *Bioinformatics*. 2015; **31**: 94-101.

18 Dierynck I, Thys K, Ghys A, et al. Deep-sequencing analysis of the gene encoding the hepatitis c virus nonstructural 3-4a protease confirms a low prevalence of telaprevir-resistant variants at baseline and the end of the realize study. *Journal of Infectious Diseases*. 2014; **210**: 1871-1880.

19 Beloukas A, Magiorkinis E, Magiorkinis G, et al. Assessment of phylogenetic sensitivity for reconstructing hiv-1 epidemiological relationships. *Virus Res*. 2012; **166**: 54-60.

20 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of phyml 3.0. *Systematic Biology*. 2010; **59**: 307-321.

21 Price MN, Dehal PS, Arkin AP. Fasttree 2--approximately maximum-likelihood trees for large alignments. *PloS One*. 2010; **5**: e9490.

22 Gianella S, Delport W, Pacold ME, et al. Detection of minority resistance during early hiv-1 infection: Natural variation and spurious detection rather than transmission and evolution of multiple viral variants. *Journal of Virology*. 2011; **85**: 8359-8367.

23 McCloskey RM, Liang RH, Joy JB, et al. Global origin and transmission of hepatitis c virus nonstructural protein 3 q80k polymorphism. *Journal of Infectious Diseases*. 2014.

24 Pickett BE, Striker R, Lefkowitz EJ. Evidence for separation of hcv subtype 1a into two distinct clades. *Journal of Viral Hepatitis*. 2011; **18**: 608-618.

25 **Lawitz E, Sulkowski MS, Ghalib R, et al. Simeprevir plus sofosbuvir, with or without ribavirin, to treat chronic infection with hepatitis c virus genotype 1 in non-responders to pegylated interferon and ribavirin and treatment-naive patients: The cosmos randomised study. *Lancet*. 2014; 384: 1756-1765.**

26 Lawitz E, Matusow G, DeJesus E. A phase 3, open-label, single-arm study to evaluate the efficacy and safety of 12 weeks of simeprevir (smv) plus sofosbuvir (sof) in treatment-naive or experienced patients with chronic HCV genotype 1 (HCV-1) infection and cirrhosis: Optimist-2. *50th Annual Meeting of the European Association for the Study of the Liver*. Vienna, Austria - April 22-26, 2015.

27 McPhee F, Friborg J, Levine S, et al. Resistance analysis of the hepatitis c virus NS3 protease inhibitor asunaprevir. *Antimicrobial Agents and Chemotherapy*. 2012; **56**: 3670-3681.

28 Pilot-Matias T, Tripathi R, Cohen D, et al. In vitro and in vivo antiviral activity and resistance profile of the hepatitis c virus NS3/4a protease inhibitor abt-450. *Antimicrobial Agents and Chemotherapy*. 2015; **59**: 988-997.

29 Lenz O, Verbinnen T, Fevery B, et al. Virology analyses of HCV isolates from genotype 1-infected patients treated with simeprevir plus peginterferon/ribavirin in phase iib/iii studies. *Journal of Hepatology*. 2014.

30 Akuta N, Suzuki F, Sezaki H, et al. Evolution of simeprevir-resistant variants over time by ultra-deep sequencing in hcv genotype 1b. *Journal of Medical Virology*. 2014; **86**: 1314-1322.

AF009606
AY695436
AF011751
AF011752
AF011753
AY615798
AY695437
AY956463
AY956464
AY956465
AY956466
AY956468
AY956469
DQ068178
DQ068179
DQ068180
DQ068181
DQ068182
DQ068183
DQ068184
DQ068185
DQ068186
DQ068187
DQ068188
DQ068189
DQ068190
DQ068191
DQ068192
DQ068193
DQ068194
DQ068195
DQ068196
DQ068197
DQ068198
DQ068199
DQ068200
DQ068201
DQ068202
DQ068203
DQ068204
DQ068205
DQ068206
DQ068207
DQ430811
DQ430812
DQ430813
DQ430814
DQ838739
DQ838740
DQ838741

DQ838742
DQ838743
DQ838744
DQ838745
DQ889251
DQ889252
DQ889253
DQ889254
DQ889255
DQ889256
DQ889257
DQ889258
DQ889259
DQ889260
DQ889261
DQ889262
DQ889263
DQ889264
DQ889265
DQ889266
DQ889267
DQ889268
DQ889269
DQ889270
DQ889271
DQ889272
DQ889273
DQ889274
DQ889275
DQ889276
DQ889277
DQ889278
DQ889279
DQ889280
DQ889281
DQ889282
DQ889283
DQ889284
DQ889286
DQ889287
DQ889288
DQ889289
DQ889290
DQ889291
DQ889292
DQ889293
DQ889294
DQ889295
DQ889296
DQ889297

DQ889298
DQ889299
DQ889300
DQ889301
DQ889302
DQ889303
DQ889304
DQ889305
DQ889306
DQ889307
DQ889308
DQ889309
DQ889310
DQ889311
DQ889312
DQ889313
DQ889314
DQ889315
DQ889316
DQ889317
DQ889318
DQ889319
DQ889320
DQ889321
EF032883
EF032884
EF032885
EF139707
EF139710
EF139715
EF139719
EF139723
EF139725
EF139728
EF139729
EF139730
EF139735
EF139736
EF139740
EF139741
EF139746
EF139748
EF139749
EF139754
EF173627
EF173628
EF173629
EF173630
EF173631
EF173632

EF173633
EF173634
EF173635
EF173636
EF173637
EF173638
EF173639
EF173640
EF173641
EF173642
EF173643
EF173644
EF173645
EF173646
EF173647
EF173648
EF173649
EF173650
EF173651
EF173652
EF173653
EF173654
EF407411
EF407412
EF407413
EF407414
EF407415
EF407416
EF407417
EF407418
EF407419
EF407420
EF407421
EF407422
EF407423
EF407424
EF407425
EF407426
EF407427
EF407428
EF407429
EF407430
EF407431
EF407432
EF407433
EF407434
EF407435
EF407436
EF407437
EF407438

EF407439
EF407440
EF407441
EF407442
EF407443
EF407444
EF407445
EF407446
EF407447
EF407448
EF407449
EF407450
EF407451
EF407452
EF407453
EF407454
EF407455
EF407456
EF407457
EF621489
EU155213
EU155214
EU155215
EU155216
EU155233
EU155236
EU155237
EU155238
EU155239
EU155240
EU155241
EU155242
EU155243
EU155244
EU155245
EU155246
EU155247
EU155248
EU155249
EU155250
EU155251
EU155252
EU155265
EU155266
EU155267
EU155268
EU155269
EU155270
EU155271
EU155272

EU155273
EU155274
EU155275
EU155276
EU155277
EU155278
EU155282
EU155283
EU155284
EU155285
EU155286
EU155287
EU155288
EU155289
EU155290
EU155291
EU155292
EU155293
EU155294
EU155295
EU155296
EU155297
EU155298
EU155299
EU155309
EU155310
EU155311
EU155312
EU155313
EU155314
EU155319
EU155320
EU155321
EU155322
EU155323
EU155338
EU155339
EU155340
EU155341
EU155342
EU155343
EU155344
EU155345
EU155346
EU155347
EU155348
EU155349
EU155350
EU155351
EU155352

EU155353
EU155354
EU155355
EU155378
EU155379
EU155380
EU234063
EU234064
EU234065
EU239713
EU239715
EU239716
EU250017
EU255927
EU255928
EU255929
EU255930
EU255931
EU255932
EU255933
EU255934
EU255935
EU255936
EU255937
EU255938
EU255939
EU255940
EU255941
EU255942
EU255943
EU255944
EU255945
EU255946
EU255947
EU255948
EU255949
EU255950
EU255951
EU255952
EU255953
EU255954
EU255955
EU255956
EU255957
EU255958
EU255959
EU255963
EU255964
EU255965
EU255966

EU255967
EU255968
EU255969
EU255970
EU255971
EU255973
EU255974
EU255975
EU255976
EU255977
EU255978
EU255979
EU255980
EU255981
EU255982
EU255983
EU255984
EU255985
EU255986
EU255987
EU255988
EU255989
EU255990
EU255991
EU255992
EU255993
EU255994
EU255995
EU255996
EU255997
EU255998
EU255999
EU256002
EU256003
EU256004
EU256005
EU256006
EU256007
EU256008
EU256009
EU256010
EU256011
EU256012
EU256013
EU256014
EU256015
EU256016
EU256017
EU256018
EU256019

EU256020
EU256021
EU256022
EU256023
EU256024
EU256025
EU256026
EU256027
EU256028
EU256029
EU256030
EU256031
EU256032
EU256033
EU256034
EU256035
EU256036
EU256037
EU256038
EU256039
EU256040
EU256041
EU256042
EU256043
EU256044
EU256046
EU256047
EU256048
EU256049
EU256050
EU256051
EU256052
EU256053
EU256055
EU256056
EU256057
EU256058
EU256060
EU256067
EU256068
EU256069
EU256070
EU256071
EU256072
EU256073
EU256074
EU256086
EU256087
EU256094
EU256095

EU256096
EU256097
EU256104
EU256105
EU256106
EU256107
EU260395
EU260396
EU362876
EU362877
EU362878
EU362879
EU362880
EU362881
EU362882
EU362883
EU362884
EU362885
EU362886
EU362887
EU362888
EU362889
EU362890
EU362891
EU362892
EU362893
EU362894
EU362895
EU362896
EU362897
EU362898
EU362899
EU362900
EU362901
EU362902
EU362903
EU362904
EU362905
EU362906
EU362907
EU362908
EU362909
EU362910
EU362911
EU482831
EU482832
EU482834
EU482835
EU482836
EU482837

EU482838
EU482840
EU482841
EU482842
EU482843
EU482845
EU482846
EU482847
EU482848
EU482850
EU482852
EU482853
EU482854
EU482855
EU482856
EU482857
EU482858
EU482861
EU482862
EU482863
EU482864
EU482865
EU482866
EU482867
EU482868
EU482869
EU482870
EU482871
EU482872
EU482873
EU482876
EU482878
EU482882
EU482884
EU482887
EU482889
EU529676
EU529677
EU529678
EU529679
EU529680
EU529681
EU569722
EU569723
EU595697
EU595698
EU595699
EU622930
EU660383
EU660384

EU660385
EU660387
EU677247
EU677248
EU677249
EU677250
EU677251
EU677252
EU677253
EU677254
EU677255
EU677256
EU677257
EU677258
EU687193
EU687194
EU687195
EU781746
EU781747
EU781748
EU781749
EU781750
EU781751
EU781752
EU781753
EU781754
EU781755
EU781756
EU781757
EU781758
EU781759
EU781760
EU781761
EU781762
EU781763
EU781764
EU781765
EU781766
EU781767
EU781768
EU781769
EU781770
EU781771
EU781772
EU781773
EU781774
EU781775
EU781776
EU781777
EU781778

EU781779
EU781781
EU781782
EU781783
EU781784
EU781785
EU781786
EU781787
EU781788
EU781789
EU781790
EU781791
EU781792
EU781793
EU781794
EU781795
EU781796
EU781797
EU781798
EU781799
EU781800
EU781801
EU781802
EU781803
EU781804
EU862823
EU862824
EU862826
EU862827
EU862828
EU862829
EU862830
EU862831
EU862832
EU862833
EU862834
EU862836
EU862838
EU862839
EU862840
EU862841
FJ024087
FJ024274
FJ024275
FJ024276
FJ024278
FJ024280
FJ024281
FJ024282
FJ181999

FJ182000
FJ182001
FJ205867
FJ205868
FJ205869
FJ390394
FJ390395
FJ390399
FJ410172
GQ149768
HM568417
HM568418
HM568419
HM568420
HM568421
HM568422
HM568423
HM568424
HM568425
HM746870
HM746873
HM746876
HM746877
HM746878
HM746879
HM746880
HM746882
JN704193
JN704194
JN704198
JN704199
JN704202
JN704203
JN704204
JN704206
JN704208
JN704210
JN704218
JN704219
JN704220
JN704224
JN704229
JN704230
JN704231
JN704232
JN704233
JN704234
JN704235
JN704236
JN704237

JN704238
JN704239
JN704240
JN704241
JN704242
JN704243
JN704244
JN704245
JN704246
JN704247
JN704248
JN704249
JN704250
JN704251
JN704252
JN704253
JN704254
JN704255
JN704256
JN704257
JN704260
JN704261
JN704262
JN704263
JN704264
JN704266
JN704267
JN704268
JN704269
JN704270
JN704273
JN704274
JN704275
JN704276
JN704277
JN704278
JN704280
JN704281
JN704282
JN704283
JN704284
JN704285
JN704290
JN704291
JN704293
JN704294
JN704295
JQ061779
JQ061780
JQ061781

JQ061782
JQ061783
JQ061784
JQ061785
JQ061786
JQ061787
JQ914271
JQ914273
JQ914274
JX463525
JX463526
JX463527
JX463528
JX463529
JX463530
JX463531
JX463532
JX463533
JX463534
JX463535
JX463536
JX463537
JX463538
JX463539
JX463540
JX463541
JX463542
JX463543
JX463544
JX463545
JX463546
JX463547
JX463548
JX463549
JX463550
JX463551
JX463552
JX463553
JX463554
JX463555
JX463556
JX463557
JX463558
JX463559
JX463560
JX463561
JX463562
JX463563
JX463564
JX463565

JX463566
JX463567
JX463568
JX463569
JX463570
JX463571
JX463572
JX463573
JX463574
JX463575
JX463576
JX463577
JX463578
JX463579
JX463580
JX463581
JX463582
JX463583
JX463584
JX463585
JX463586
JX463587
JX463588
JX463589
JX463590
JX463591
JX463592
JX463593
JX463594
JX463595
JX463596
JX463597
JX463598
JX463599
JX463600
JX463601
JX463602
JX463603
JX463604
JX463605
JX463606
JX463607
JX463608
JX463609
JX463610
JX463611
JX463612
JX463613
JX463614
JX463615

JX463616
JX463617
JX463618
JX463619
JX463620
JX463621
JX463622
JX463623
JX463624
JX463625
JX463626
JX463627
JX463628
JX463629
JX463630
JX463631
JX463632
JX463633
JX463634
JX463635
JX463636
JX463637
JX463638
JX463639
JX463640
JX463641
KC155254
M62321
M67463
NC_004102